

ОТЗЫВ

отечественного консультанта о диссертационном исследовании
ГОЛЕНКО ЕКАТЕРИНЫ СЕРГЕЕВНОЙ
«Разработка алгоритмов анализа данных масс-спектрометрии
нативных белков», представленном на соискание степени
доктора философии (PhD) по образовательной программе
«8D06101 - Аналитика больших данных»

Диссертация Голенко Екатерины Сергеевны на тему «Разработка алгоритмов анализа данных масс-спектрометрии нативных белков» посвящена актуальной проблеме. В настоящее время масс-спектрометрия является одним из самых распространенных методов в исследованиях пептидов и белков. Быстрое развитие данного метода способствует увеличению скорости, чувствительности, качества данных и надежности результатов. Появляются новые инструменты, повышаются производительность и эффективность программного обеспечения, скорость получения данных и удобство использования. При этом исследователи получают гигабайты данных в течение нескольких часов, что затрудняет их анализ вручную. Таким образом, обработка экспериментальных данных в исследованиях пептидов и белков представляет собой сложную задачу, для решения которой требуются обширные знания многих пакетов программ, имеющих различные алгоритмы, требования к формату данных и пользовательские интерфейсы.

Множество факторов, включая присутствие посттрансляционных модификаций, разложение белков на фрагменты и проблемы с ионизацией, влияют на то, что только ограниченный набор белков может быть точно идентифицирован. Современные методы обычно позволяют достоверно идентифицировать менее половины всех белков в образце, оставляя значительное количество потенциально важных молекул без подтверждения их присутствия и функции. Это создаёт существенный барьер для понимания сложных биологических систем и требует дальнейшего усовершенствования методов масс-спектрометрии и разработки более мощных алгоритмов для обработки и интерпретации спектральных данных.

Материалами для исследования послужили: 1. Данные масс-спектрометрии, полученные в рамках проекта «ПЦР-тест для детекции и дифференциальной диагностики возбудителей описторхоза и меторхоза» под руководством Киян В.С., PhD, Научно-исследовательская платформа сельскохозяйственной биотехнологии (НИПСБ).

2. Данные из баз данных белковых и ДНК-последовательностей NIST, PRIDE, Pfam, GenBank и UniProtKB/Swiss-Prot.

В своей работе Голенко Е.С. предложила модели и методы, которые позволяют обрабатывать данные масс-спектрометрии, одновременно используя как машинное обучение, так и методы сравнения последовательностей.

В ходе проведения диссертационного исследования Голенко Е.С. показала отличную способность формулировать собственную точку зрения по рассматриваемой проблеме. Сформулированные в работе выводы достаточно обоснованы, друг другу не противоречат, и могут быть использованы в практической деятельности.

В диссертации получены следующие результаты:

1. Разработан алгоритм для идентификации пептидов, разработанный на основе сети подобию с открытым исходным кодом SpeCollate, использующий нейронную сеть BiLSTM для поиска совпадений пептидного спектра.

2. Разработан алгоритм аннотации белковых функций, построенный на основе комбинации нейронной сети BiLSTM и механизма самовнимания (self-attention).

Предложенные алгоритмы способны обрабатывать большие объемы данных и интегрировать информацию на множественных уровнях, включая генетическую, метаболическую и фенотипическую информацию.

Практическая значимость результатов диссертационного исследования заключается в применимости результатов и имплементации их в программные модули лабораторий для идентификации белковых последовательностей и предсказания их функций с высокой надежностью.

Научно-обоснованные теоретические и экспериментальные результаты диссертационной работы использованы в научном проекте по теме «ПЦР-тест для детекции и дифференциальной диагностики возбудителей описторхоза и меторхоза».

Личный вклад автора состоит в непосредственном выполнении исследований по всем главам и логическим звеньям диссертации: проведение обзора и анализа ранее представленных работ, выбор и обоснование использованных методов, разработка и техническая реализация алгоритмов, апробация и тестирование разработанных моделей на исходных данных. Соискателем материал диссертационной работы изложен литературно, логически последовательно и оформлен, в соответствии с установленными требованиями к подобным работам.

В ходе диссертационного исследования автором было опубликовано 13 научных трудов, из них 1 статья в научном журнале с ненулевым импакт-фактором, входящим в международную базу SCOPUS (процентиль по CiteScore2022 равный 34), 3 статьи в журналах, рекомендованных Комитетом по обеспечению качества в сфере науки и высшего образования Министерства науки и высшего образования Республики Казахстан, 6 статей в сборниках международных конференций, 3 статьи – в других изданиях.

Голенко Екатерина прошла научную стажировку в 2022 году на базе Института систем информатики им. А.П. Ершова Сибирского отделения Российской академии наук, РФ, г. Новосибирск под руководством зарубежного консультанта, к.ф.-м.н. Штокало Д.Н.

На основании вышеизложенного, считаю, диссертационную работу Голенко Екатерины Сергеевны на тему «Разработка алгоритмов анализа данных масс-спектрометрии нативных белков» завершённой, удовлетворяющей требованиям, предъявляемым к диссертациям на соискание доктора философии (PhD) по образовательной программе «8D06101 - Аналитика больших данных» и рекомендую к публичной защите.

Научный консультант
PhD, ассоциированный профессор,
и.о. профессора кафедры
«Информационные системы»
Казахского агротехнического
исследовательского университета
им. С.Сейфуллина



Исмаилова А.А.

19.01.2024